Information Theory and All That Jazz:
A Lost Reference List Leads to a Pragmatic
Assignment for Students

Number 44, October 31, 1977

In 1948 Claude Shannon announced his mathematical theory of communication,[1] which has since found wide applications in fields as diverse as electronic engineering and the life sciences. Shannon's mathematical model of communications systems has been widely adopted both by information theorists, who are concerned with the fundamental limitations in the transmission of information, and by communications theorists, who mainly are concerned with the operation of communications devices.

When I first entered the field of documentation—now called "information science"—in 1951, it was quite fashionable to discuss Shannon's theory.[2] In the course of the next decade much was said about information theory—but little of practical use was accomplished in the area of information retrieval.

Back in those days, information people were preoccupied with punched cards. The bible of early documentalists was Casey and Perry's book, *Punched Cards: Their Application to Science and Industry.*[3] It now amazes me to look back and realize that the punched card has already become obsolete as far as the information science literature is concerned. You certainly would never realize from reading professional information science journals that the punched card is still very much alive in the US and elsewhere.

I was raised and nurtured on the punched card. But by the mid-50s I was already using direct keyboard-tape input in Univac I—the first commercial electronic digital computer to use magnetic tape. I came to know the Univac well, not only through my acquaintance with John Mauchly but also because I used the machine for my doctoral dissertation.[4]

In 1959 I presented a paper on the role of information theory in the design of punched-card systems to a

271

meeting of the American Documentation Institute. The main purpose of the paper was to show how little relevance information theory had to information retrieval. Information science and information theory are quite distinct fields.

After my presentation, Herman Skolnik, who later became the first editor of the *Journal of Chemical Documentation* (now called the *Journal of Chemical Information and Computer Science*), asked me for a copy of my talk. The only copy I had was the one I'd just finished reading. Since I was so flattered by his request I gave it to him immediately. Unfortunately, the numerous references I cited were not attached or indicated in the text. Little did I realize that two years later the paper would be published verbatim—and without the missing references. Thus the foremost proponent of citation indexing became the author of a review paper without a single cited reference. To make matters worse, I lost the list of documents cited. So I never submitted a suitable correction.

The original list of references is still missing. However, I recently reconstructed it. The appropriate references are included here in the reprint of the original paper.

The loss of the references was a fortuitous blessing. As an adjunct lecturer at the University of Pennsylvania, I have used the paper as the basis of an interesting teaching exercise in documentation. In order to illustrate the variable nature of citation behavior, I have asked my students to read the paper carefully. Each student then inserts parentheses wherever he or she feels that a reference is needed. I have now performed this exercise for more than ten years. Each time the result has been the same. Although the number of references students inserted varied from about fifteen to seventy-five, they consistently averaged about forty-five. That is just about the number my original manuscript contained. In fact, the reconstructed reference list contains 41 items.

I'm convinced that many more instructors should use assignments of this type to train students in literature searching and citation consciousness. Developing a basic feel for when a reference is appropriate as well as skill in tracking down an elusive citation will serve them well during their professional careers.

Although part of the article which follows may now seem obsolete, it should be clear to most readers that the punched card remains a ubiquitous fact of life. Anyone who has ever received a government check, a phone bill, or any one of numerous computer-prepared notices should not dispute this.

But apart from the continued

relevance of punched-card technology in our lives, I cannot resist the opportunity to debunk, indirectly, the unnecessary mathematical gibberish which appears not only in journals of information science but also in the literature of other disciplines. A few years ago my old professor of chemistry, Joel Hildebrand, complained about this very problem in chemical physics.[5]

In closing, I would remind *CC®* readers that ISI® 's director of research, A.E. Cawkell, has published some highly readable articles on information theory from the viewpoint of the information engineer.[6-8]

REFERENCES

1. **Shannon C E.** A mathematical theory of communication. *Bell System Technical Journal* 27:379, 623, 1948.
2. **Mooers C N.** Coding, information retrieval and the rapid selector. *American Documentation* 1:225, 1950.
3. **Casey R S, Perry J W, Berry M M & Kent A,** eds. *Punched cards: their application to science and industry.* New York: Reinhold, 1958.
4. **Garfield E.** Chemico-linguistics: computer translation of chemical nomenclature. *Nature* 192:192, 1961.
5. **Hildebrand J.** Operations on swollen theories with Occam's razor. *Structure-solubility relationships in polymers.* (Harris F W & Seymore R B, eds.) New York: Academic Press, 1977, p. 1-9.
6. **Cawkell A E.** Simplified information theory and data transmission. I. *Electrical Engineering* 39:212, 1967.
7. ——————. Simplified information theory and data transmission. II. *Electrical Engineering* 39:302, 1967.
8. ——————. A measure of 'efficiency factor' — communication theory applied to document selection systems. *Information Processing & Management* 11:243, 1975.